Powder Pattern Prediction

Armel Le Bail – Université du Maine, Laboratoire des Fluorures, CNRS UMR 6010, Avenue O. Messiaen, 72085 Le Mans, Cedex 9 – alb@cristal.org

Powder patterns are predictable by a direct consequence of crystal structure predictability. The ICDD PDF-2 contains already a few hundreds of theoretical powder patterns coming from special ICSD data (having the THE comment - for theoretical) used for building their calculated powder patterns. For instance there are 24 theoretical SiO_2 powder patterns in PDF-2, calculated from the atomic coordinates coming from ref [1] and deposited in ICSD. As a user I would prefer to know if the data from a database are experimental or not. But I surely would be interested in saving time by identifying my new compound as having been predicted already.

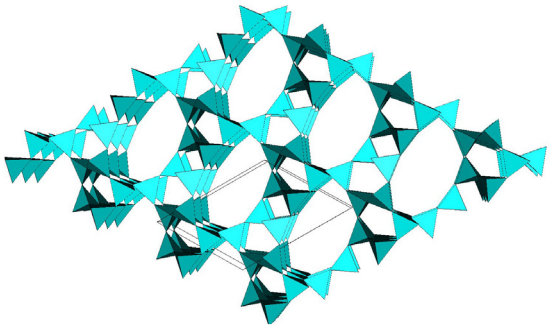

Fig. 1 - Will that predicted structure (PCOD1050015) be really observed some day ?

"Are structure predictable?" was the title of a 1994 publication [2]. Ten years after, considerable progress have been made in structure prediction, enumeration, simulation, polymorph-generation, etc, either by using force-field, energy-minimization, distance-least-squares, valence-rule, simulated-annealing, design, packing-considerations, etc. The most active field seems to be in "protein structure prediction" nowadays, a Google search with these three words returns 220000 pages while "inorganic structure prediction" returns 28000, most of the latter pointing at a few commercial software. Thousands of structures have been predicted already, unfortunately they cannot be retrieved very simply and searched. Moreover, there are very few chances that ICDD/ICSD/CSD/CRYSTMET would accept the massive deposition of purely theoretical powder patterns or crystal data. My opinion is that the small and medium predicted crystal structures need for a special database (let the proteins apart). Software for phase identification by search-match in a powder pattern database give the choice for selecting some data subsets : mineral, organic (etc) and so, why not a theoretical subset ?

Being mostly interested in inorganic compounds, I considered that making a database of predicted compounds would be better than waiting indefinitely for it. First of all, a database must be fed. I decided to become myself a predictor, starting by the most simple approach, using geometrical restraints applied to some series of well defined polyhedra-based compounds (tetrahedra, octahedra...). This concept is of course absolutely not new. Zeolite researchers have documented more than 1000 hypothetical structures by using classical physical model building [3] during the past 60 years. Simulated annealing is a rapid generator of hypothetical 4-connected framework structures and others. More than 5000 hypothetical zeolite structures were reported in ref. [4]. Many recent works in inorganic structure prediction have produced huge quantities of hypothetical compounds (using commercial packages as CERIUS, etc), there is no room here for citing them all. However, no systematic categorization was attempted so that, if interested, people would have to rebuild these hypothetical structures by themselves.

A preliminary version of the predicting sofware is available, named GRINSP (Geometrically Restrained INorganic Structure Prediction) [5]. The main purpose of GRINSP is to generate hypothetical structures which will be documented in a searchable database : PCOD (Predicted Crystallography Open Database), a subset of the COD [6]. If you are a structure predictor and want to deposit your predicted structures in PCOD, that is already possible (CIF files only).

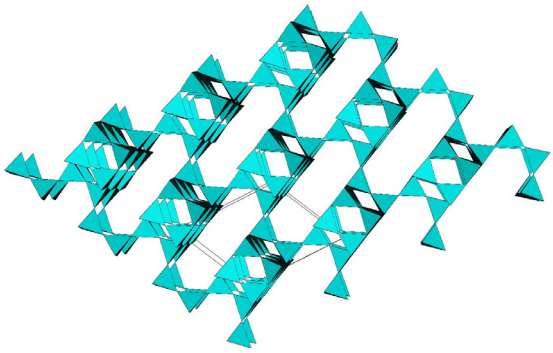

Fig. 2 - Another improbable zeolite : PCOD1050017.

GRINSP does not work by applying simulated annealing to a starting random configuration. Version 0.9 works schematically as follows, by using the Monte Carlo method : manual selection of the constraints on cell parameters and of restrained interatomic distances; random selection of the cell parameters inside of the predefined range; random positioning of a first T atom of the TX_2 compound; random positioning of the next T atoms in respect of the distance restraints with the previous ones; if a model fulfills all criteria, place the X atoms at T-T midpoints, refine the atomic positions and cell parameters; continue to try to predict structures in that way till a certain number of independent runs are made; find if the predicted structures are new or were already described (using the CS fingerprint - Coordination Sequence).

In the GRINSP algorithm, the number of T atoms in a randomly selected cell is not predetermined, it is predicted as well. Only distances are considered (not angles - though considering a range for the second T-T distances is like giving angles). Currently, there are many limitations in that preliminary beta version. The choice of the space group is limited to P1. So that, GRINSP 0.9 is only efficient for a maximum number of 10-20 T atoms in a triclinic cell (which you may force to be cubic-like or hexagonal-like or etc). It will be able to predict only the very small zeolites (ABW, EDI, BIK...) or the compact SiO_2 phases (quartz, etc) and also a lot of hitherto unknown phases (see the PCOD [6]). Further work is thus needed for improving the GRINSP efficiency: introduction of all space groups and management of special positions (considering only the unique T atoms); introduction of other polyhedra types (octahedra in a first time), predicting structures combining them; increase speed by not recalculating always everything (distances); Increase the box size for the CS (coordination sequence) calculations (729 cells is not always enough); reduce the output to the non-redondant predicted structures; etc ! GRINSP is distributed under the GNU Public License - so, you may decide to make improvements by yourself, provided the modified source code is made available under the same licence.

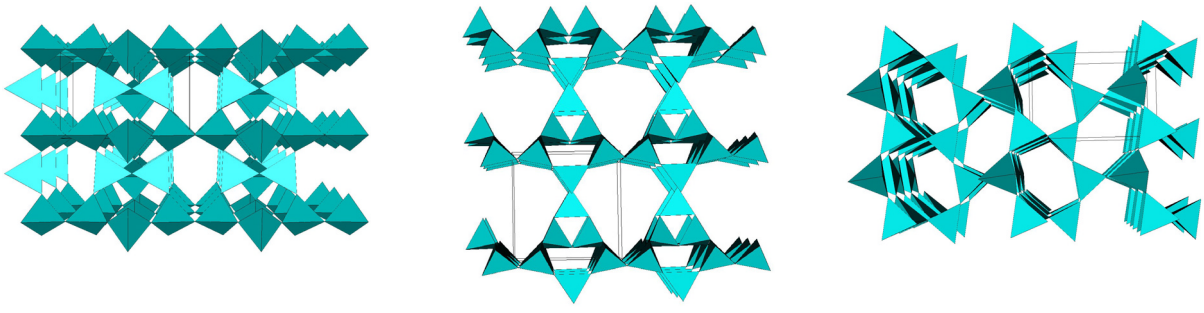

Fig. 3 - Three orthogonal views of PCOD1020007.

It can be anticipated that predicted/enumerated crystal structures may soon (during the next decades) exceed in number the really determined ones. It is probable that the "one stucture - one publication" which is almost the rule for the real crystal data will never be applied for the virtual ones. Therefore, it is logical to offer room for storage of these predicted or enumerated crystal structures. As a beginning, PCOD [6] contains an enumeration of 150 SiO_2 polymorphs built up by using GRINSP. Predicting software should propose in first place in their lists the really observed compounds. This is the case of GRINSP which always places quartz, tridymite, cristobalite, EDI, ABW, BIK, JBW, SOD, YUG, etc, at the first places of the smallest predicted cells, with parameters close to the real ones. The next step for PCOD, if high standards are attained in the quality of the predicted models, is to produce a set of predicted powder patterns usable by search-match software. Thanks to the PCOD, phase identification would be extended to the hitherto unknown compounds, saving time in difficult structure determinations. But that dream is far from being realized yet, in spite of thousands of predicted crystal structures listed in the past 20 years (including thousands of zeolites).

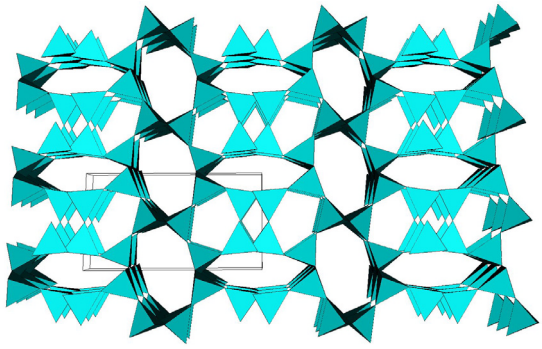

Fig. 4 - The predicted and very improbable PCOD1050025.

Only cell parameters in the range 4.9-13 A were considered in the preliminary predictions. Other known zeolites "predicted" in this range are triclinic or monoclinic BIK, SOD and YUG.

Even if many chimeric compounds will be added into PCOD, we can expect some treasures too. Either chimeric or to-be-observed-some-day, some structures as predicted by GRINSP are at least nice and decorative (see the figures). And one can expect that the real compounds (quartz, cristobalite, etc), retrieved and simulated by GRINSP as well, would be identified by a search-match from theoretical powder patterns built up from the data listed in the Table I. So that powder pattern prediction is really the next step. When PCOD will have a non-negligible number of entries, a file of *d*s and *I*s will be built up automatically from the crystal data (CIF files) and proposed for compilation to the search-match software. Why not to expect even some day that predictions will be as accurate as structures really determined and refined. That day, a database of real structures will have much less interest. About redundancy, and prediction of variants of topologically identical phases, should they all be inserted into the PCOD ? I would say yes, since databases of real compounds are also plenty of such variants.

Table I - Comparison of some GRINSP-predicted cell parameters with dense SiO_2 polymorphs or zeolites, idealized or observed ones :

	Predicted / *Observed*			
	a	b	c	$R_{(DLS)}$
Quartz	4.965	4.965	5.375	0.0009
	4.912	*4.912*	*5.404*	
Tridymite	5.073	5.073	8.400	0.0045
	5.052	*5.052*	*8.270*	
Cristobalite	5.024	5.024	6.796	0.0018
	4.969	*4.969*	*6.926*	
ABW	9.331	5.175	8.873	0.0068
	9.9	*5.3*	*8.8*	
EDI	6.919	6.919	6.407	0.0047
	6.926	*6.926*	*6.410*	
JBW	5.209	7.543	7.983	0.0066
	5.3	*7.5*	*8.2*	

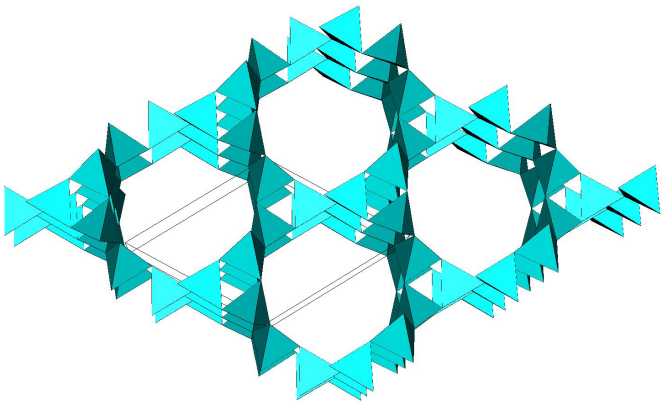

Fig. 5 - PCOD1030007 with large ring aperture, another chimeric zeolite.

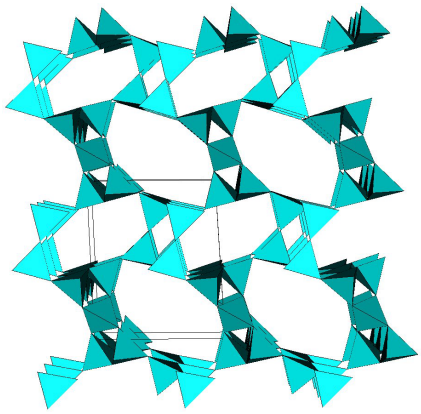

Fig. 6 - PCOD1060041, improbable triclinic simulation, or not?

So, do not only mention and show nice pictures of your huge quantities of predicted crystal structures in your manuscripts, also deposit their atomic coordinates at PCOD, as CIF files. Thanks !

[1] M.B. jr Boisen, G.V. Gibbs, M.S.T. Bukowinski, Phys. & Chem. of Minerals (Germany) 21 (1994) 269-284.
[2] A. Gavezzotti, Acc. Chem. Res. 27 (1994) 309-314.
[3] J.V. Smith, Chem. Rev. 88 (1988) 149-182.
[4] M.W. Deem and J.M. Newsam, J. Am. Chem. Soc. 114 (1992) 7189-7198.
[5] A. Le Bail (2003), Program GRINSP : http://sdpd.univ-lemans.fr/grinsp/
[6] COD (Crystallography Open Database) : http://www.crystallography.net/